Dorian: Grid Service Infrastructure for Identity Management and Federation

Stephen Langella[1], Scott Oster[1], Shannon Hastings[1], Frank Siebenlist[2], Tahsin Kurc[1], Joel Saltz[1]

[1]*Department of Biomedical Informatics*
Ohio State University
Columbus, OH 43210
{langella,oster,hastings,kurc,saltz}@bmi.osu.edu

[2]*Mathematics and Computer Science Division*
Argonne National Laboratory
Argonne, IL 60439
franks@mcs.anl.gov

Abstract

Identity management and federation is becoming an ever present problem in large multi-institutional environments. By their nature, Grids span multiple institutional administration boundaries and aim to provide support for the sharing of applications, data, and computational resources in a collaborative environment. One underlying problem is to enable participating institutions to manage the identities of their own members by leveraging existing institutional identity management systems, while at the same time facilitating the participation in larger Grids through the deployment of grid-wide user credentials. Those grid-wide identities are used for features such as single sign-on, secure communication, and are the basis for authorization decisions. In this paper we will present the design and implementation of Dorian, a grid service infrastructure component that enables the federation of users across the collaboration.

1. Introduction

As Grid technologies gain acceptance and adoption, the transition from highly specialized Grids with only a few institutional participants to more general purpose Grids with dozens to hundreds of institutions is becoming a reality. Security is of primary importance in such environments, in particular in settings where sensitive data (e.g., patient medical information) needs to be accessed and exchanged. Mechanisms are needed for such functions as secure communication, authentication, and authorization. This paper is concerned with the management and federation of user identities and authentication in the Grid.

Identity management and federation is becoming a critical problem in enabling access to resources across multiple security domains. In such an environment, it is highly unlikely that participating institutions use the same mechanism for managing the identities of their members. A common application of identity federation in the business domain is to enable single sign-on between two interacting companies that employ different internal authentication mechanisms, and vouch for their access to resources without needing to adopt the same security technologies or maintain a shared, centralized system for managing member identities. The Grid introduces additional challenges since it aims to facilitate the sharing of resources in a collaborative environment that spans multiple institutional boundaries. As a consequence, a Grid environment has to facilitate access by its users to resources at disparate institutions, while enforcing authentication and authorization policies set forth by the different institutions.

When the number of institutions participating in a Grid environment reaches hundreds, the number of individuals wishing to join the Grid can be expected to be well within the neighborhood of thousands or tens of thousands. This creates a huge challenge in terms of managing users, creating and managing their Grid identities, and supporting authentication such that Grid services can be accessed securely. When the number of users is large, it would be neither desirable nor tractable to have a central system to manage the identities for all participants. It is also undesirable to require that each shared service or group of services in the Grid maintains separate identity management and authentication mechanisms. Therefore, a Grid-enabled infrastructure is needed to facilitate the federation of institutional identities. In this paper we describe the design and implementation of Dorian, a Grid service infrastructure to address these issues.

Dorian[1] provides a complete Grid-enabled solution, based on public key certificates and SAML, for managing and federating user identities in a Grid environment. Grid technologies have adopted the use of X509 identity certificates to support user authentication. The Security Assertion Markup Language (SAML) [9] has been developed as a standard for exchanging authentication and authorization statements between security domains. Note that Grid certificates and SAML assertions serve different purposes. SAML is mainly used between institutions for securely exchanging authentication

[1] This work was supported in part by the National Cancer Institute's Center for Biomedical Informatics, under the caBIG™ project and in part by the Mathematical, Information, and Computational Sciences Division subprogram of the Office of Advanced Scientific Computing Research, U.S. Dept. of Energy, under Contract W-31-109-Eng-38

information coming from trusted identity providers. The primary use of the certificates is to uniquely identify Grid users, facilitate authentication and authorization across multiple resource providers, and to enable secure delegation of credentials such that a service or a client program can access resources on behalf of the user. A salient feature of Dorian is that it provides a mechanism for the combined use of both SAML [9] and Grid certificates to authenticate users to the Grid environment through their institution's authentication mechanism. Furthermore, Dorian also hides the complexities of creating and managing Grid credentials from the users.

2. Background

Our work is driven mainly by use cases and requirements gathered from the Cancer Biomedical Informatics Grid (caBIG™) [1]. The caBIG™ Security Technology Evaluation White Paper [13] serves as one of the motivating influences for the Dorian work.

2.1 caBIG™ and Grid Authentication

The caBIG™ program funded by the National Cancer Institute (NCI) was launched to meet the need for a more coordinated approach to informatics resource development, management and dissemination for cancer research. The goal is to speed the delivery of innovative approaches for the prevention and treatment of cancer by facilitating discovery, integration, and analysis of distributed information and sharing of results from multiple institutions.

Given the sensitivity of the medically related data and the number of institutions involved, security has quickly become a high priority issue in caBIG™. A key security challenge in caBIG™ is to be able to implement an effective mechanism of authenticating users to caBIG™ services, such that access to resources can be restricted to individual users or a group of users. This is a challenging issue because of the scale of caBIG™. Although the caBIG™ effort is a relatively recent, it is expected that the caBIG™ community will grow to consist of hundreds of organizations and thousands of cancer-research participants from geographically dispersed medical centers, universities, government agencies, and commercial companies.

The Grid software infrastructure of caBIG™, called caGrid [2] is based on a service-oriented architecture and provides the implementation of the core services, toolkits and wizards for the development and deployment of community provided services. One of the primary design principles of caGrid is the leveraging of open Grid standards [3, 4]. The caGrid infrastructure is built on top of the Globus Toolkit [5], the most widely used reference implementation of the grid standards. The Globus Toolkit implements support for security via its Grid Security Infrastructure (GSI) [6, 7]. The GSI utilizes X509 Identity Certificates for identifying a user. An X509 Certificate with its corresponding private key forms a unique credential or so-called "grid credential" within the Grid. These grid credentials are used to authenticate both users and services. Although this approach is very effective and secure, it is difficult to manage in a multi-institutional environment. Using the base Globus toolkit, the provisioning of grid credentials is a manual process, which is far too complicated for users. The overall process is further complicated if a user wishes to authenticate from multiple locations, as a copy of their private key and certificate has to be present a every location. Not only is this process complicated, securely distributing private keys is error prone and poses a security risk.

2.2 Grid User Management Service (GUMS)

In order to reduce the complexities of managing Grid credentials, the Grid User Management Service (GUMS) [8] has been developed by the caGrid team as part of the caGrid 0.5 release. GUMS provides support for registering and managing users in a Grid environment. It simplifies identity management by managing the Grid credentials (private key, certificate) for Grid users. A rich set of user management operations are available in the GUMS infrastructure, allowing administrators to easily manage grid users. GUMS supports service-side storage and management of Grid credentials, removing the need for users to copy credentials from machine to machine. A potential grid user employs GUMS to register for a caGrid account. The user fills out a GUMS form to request caGrid registration, and to assert various caBIG™ related attributes such as institutional affiliation. It is the responsibility of a caBIG™ administrator to authenticate the user request and to decide whether to approve the account. On account approval, GUMS creates a user account and generates associated grid credentials using Globus' simpleCA for the certificate generation and management. Once an account is established, a Grid user can use GUMS to create a Grid proxy simply by supplying a username and password.

Although GUMS removes the complexities of managing grid credentials, it does not help with the complexities of bringing tens of thousands of users with existing institutional accounts to the grid. A practical solution to this problem, both from the point

of view of the users' and their institutions, is to allow those users to authenticate with the grid through the same mechanism in which they authenticate with their institution. In the next section we will present Dorian, the next generation of GUMS and a grid service infrastructure for identity mananagment and federation. Like its predecessor, Dorian hides the complexities of creating and managing grid credentials from the users. In addition, Dorian provides a mechanism for users to authenticate using their institution's authentication mechanism. This assumes that a trust agreement is in place between Dorian and the institution.

3. Dorian

One of the challenges in building an identity management and federation infrastructure is to create an architecture that incorporates multiple differing authentication mechanisms used by various institutions. In addressing this challenge we identify two possible approaches. The first is to build an infrastructure that would allow pluggable authentication modules, wherein a module would be developed for each authentication mechanism. In this architecture, a user's authentication information would be routed to the appropriate module that contains the logic for authenticating the user with its institution. Although this approach solves the problem, it requires at least one module be developed for each authentication mechanism. This would require the Grid infrastructure administrators to become intimately familiar with each institution's authentication mechanisms, and would increase the system's complexity with each new module added.

Another approach would be for the infrastructure to accept an institutionally supplied, standard "token" as a method of authentication. In this approach users would first authenticate with their institution's identity management system. Upon successfully authentication the institution's identity management system issues a token which can then be given to the federated grid identity management system in exchange for grid credentials. The benefit of this approach over the first is that it does not require writing a plug-in every time a new institutional authentication mechanism comes online. It does, however, require every institutional authentication system to agree upon and be able to provide a common token. As SAML has been adopted [9] by many institutions, we have chosen that token format as the basis of the second approach for Dorian.

The Security Assertion Markup Language (SAML) [9] is an XML standard for exchanging authentication and authorization data between security domains. Generally the exchange of authentication and authorization data is made between an *Identity Provider* (IdP) and another party. An institution's authentication system or identity management system is an example of an IdP. Dorian uses SAML authentication assertions as the enabling mechanism for federating users from local institutions to the grid.

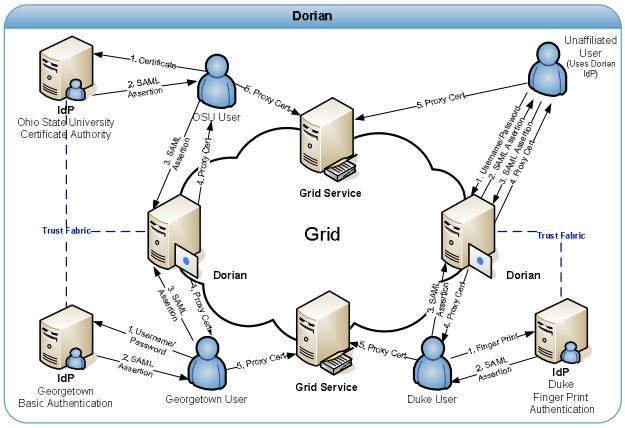

Figure 1 Example Dorian Usage Scenarios

Figure 1 illustrates an example usage scenario for Dorian. To obtain grid credentials or a proxy certificate, users authenticate with their institution using the institution's conventional mechanism. Upon successfully authenticating the user, the local institution issues a digitally signed SAML assertion, vouching that the user has authenticated. The user then sends this SAML assertion to Dorian in exchange for grid credentials. Dorian will only issue grid credentials to users that supply a SAML assertion from a *Trusted Identity Provider*. Dorian's grid service interface provides mechanisms for managing trusted identity providers; this will be discussed in greater detail later in this document. For example, in Figure 1 where a Georgetown user wishes to invoke a grid service that requires grid credentials, they first supply the application with their username and password to the Georgetown IdP as they would normally do. The application client authenticates the Georgetown user with the Georgetown IdP, receives a signed SAML assertion which it subsequently passes to Dorian in exchange for grid credentials. These credentials can then be used to invoke the grid services. This illustrates how Dorian can leverage an institution's existing authentication mechanism and bring its users to the grid.

To facilitate smaller groups or institutions without an existing IdP, Dorian also has its own internal IdP. This allows users to authenticate to Dorian directly, thereby enabling them to access the grid. It provides administrators with facilities for approving and

managing users. All of the Dorian IdP's functionality is made available through a grid service interface. Details of the Dorian IdP are provided later in this document. Figure 1 illustrates a scenario of a client using the Dorian IdP to authenticate to the Grid. In this scenario, the unaffiliated User wishes to invoke a grid service. Given that this unaffiliated user has registered and been approved for an account, she is able to authenticate with the Dorian IdP by supplying their username and password. Upon successfully authenticating the user, the Dorian IdP issues a SAML Assertion just like institutional IdPs, which can be presented to Dorian in exchange for grid credentials. The credentials can be used to invoke the grid service.

3.1. Architectural Overview

The high level architecture is illustrated in Figure 2. Dorian is built on top of the Globus Toolkit and runs as a WSRF [14] compliant grid service. Clients communicate with Dorian though its web service interface. All communication between clients and Dorian is secured via Transport Level Security (TLS) or WS-SecureConversation, depending on the deployment configuration.

The Dorian architecture consists of two core components: the Identity Federation Service (IFS) and the Dorian Identity Provider (IdP). The IFS component handles all the identity federation and management aspects for Dorian including the management of grid user accounts and Trusted IdPs. Dorian's internal IdP component provides the functionality for registering, authenticating, and managing users. Dorian provides a complete client API which provides complete programmatic access to all operations. Dorian also provides a complete graphical user interface.

3.2. Identity Federation Service (IFS)

The Identity Federation Service (IFS) component of Dorian facilitates the federation of the local user accounts from multiple institutions to the grid. Architecturally (Figure 2) the IFS consists of four components: the Trusted IdP Manager, the Certificate Authority, the Grid User Manager, and the Grid Credentials Manager. The Trusted IdP Manager component manages the list of Institutional IdPs from which Dorian will accept SAML assertions as a mechanism of authentication. The Certificate Authority component provides Dorian the ability to create and renew grid credentials for users. By default, Dorian uses its own internal certificate authority, but it can be configured to use an external certificate authority of choice. The Grid Credentials Manager component interfaces with the Certificate Authority and maintains the private key and certificate for each user. Finally, the Grid User Manager component manages the account information for each grid user. This information includes the user's identity provider, institutional user id, email address, account status, and the user's role within the IFS.

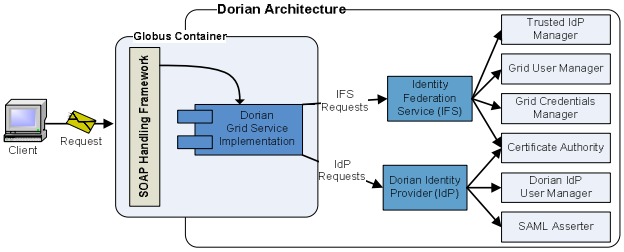

Figure 2 Dorian Components

The IFS exposes its functionality through Dorian's grid service interface. This functionality can be divided into two sub interfaces, the IFS user interface, and the IFS administrative interface. The IFS user interface allows local users to create grid credentials. The IFS administrative interface provides operations that allow administrators to manage Trusted IdPs and grid user accounts. Note that invoking an administrative operation requires a grid proxy of an administrative user.

3.2.1. Trusted Identity Providers. The institutional Identity Providers (IdPs) that Dorian is configured to trust are referred to as Trusted Identity Providers (Trusted IdPs). Dorian only creates credentials for users whose identity assertions come from a Trusted IdP. The set of Trusted IdPs can be managed by Dorian IFS administrators through its grid service interface. The Dorian grid service interface provides functionality for adding, modifying, and removing Trusted IdPs. The Trusted IdP information consists of the following: IdP Id, IdP Name, IdP Status, User Policy, Certificate, and acceptable authentication methods. The IdP Id is a unique id assigned by Dorian to identify the IdP. The IdP name is assigned by an administrator and provides human readable name to easily identify an IdP. The IdP Status specifies the current status of the IdP: Active or Suspended. Users associated with a "suspended" IdP will be refused access to Dorian. Each Trusted IdP is associated with a set of configurable User Policies that are applied to each user when they authenticate. These policies designate how Dorian should handle users from a specified Trusted IdP. As an example, a policy might dictate what to do when a new user tries to create grid

credentials for the first time. An automatic approval policy would automatically register the user with Dorian and create a grid account for the user. A manual approval policy would automatically register the user but not enable the grid account until an administrator manually approves it. User policies can also be used to dictate what to do when a user's grid credentials expire. For example, an automatic renewal policy would enable automatic creation of a new set of credentials using the Dorian certificate authority, where as a manual renewal policy would require and administrator to do so. The User Policy framework is extensible; administrators can implement local policies.

Each Trusted IdP must also specify its own certificate. When Dorian receives a SAML assertion from a Trusted IdP it verifies that the assertion was signed with the private key that corresponds to the Trusted IdP's certificate. Finally, each Trusted IdP must be configured with a list of acceptable authentication methods. A SAML authentication assertion specifies the method in which the Trusted IdP authenticated the user. In order for the SAML assertion to be accepted by Dorian, the authentication method specified in the assertion must be specified as acceptable in the corresponding Trusted IdP.

3.2.2. Grid User Management. When a user first attempts to create a proxy using Dorian, a grid user account is created for them. The account includes user information, user status, user role, and a set of grid credentials including the associated grid identity. The user information includes the user's local institution id, the id of the Trusted IdP the user is associated with, and an email address. The user's status corresponds to the user's current status: Active, Suspended, Pending, or Expired. Only users with an "Active" status may access Dorian. A user's role specifies whether or not the grid user is a Dorian IFS administrator. Only administrators may access the administrative functionality to manage Trusted IdPs or to manage grid accounts. A user's grid credentials consist of a certificate and private key that are used by Dorian to issue grid proxy certificates. A user's grid identity is compromised of the Certificate Authority's Subject DN (Distinguished Name), the IdP Id, and the user's id at his institution. When a user's grid account is created the initial status of the account is "Pending". As mentioned earlier, if the TrustedIdP has an Auto Approval User Policy in place, the status will automatically be changed to "Active", giving the user instant access to Dorian. Administrators can update a user's status and role, and can renew a user credentials.

3.2.3. Proxy Creation. Users authenticate with grid services using grid proxy certificate [15]. Such a grid "proxy" is a short-term credential (private key and certificate) that is created from a user's long-term grid credentials. Dorian facilitates the creation of grid proxies for its users. To create a grid proxy the user supplies a proxy lifetime and the SAML assertion provided by their identity provider to the Dorian client. The Dorian client generates a new public/private key pair and sends the proxy lifetime, public key, and SAML assertion to the Dorian Grid Service. The Dorian Grid Service validates the SAML assertion and employs the user's previously stored grid credentials to create and sign a proxy certificate for the user-supplied public key. The proxy certificate is then returned to the user. The proxy certificate and locally generated private key can then be used as a grid proxy credential to invoke secure grid services. It is important to note that throughout this process no sensitive information, i.e. private keys, are passed over the network.

Dorian will also provide an additional security mechanism to prevent any party other than the party that authenticated with the IdP to submit a SAML assertion to Dorian. This is accomplished by having the authenticating client create a public/private key pair. The public key is sent by the client to the IdP with their authentication request. After authenticating the client the IdP includes the public key supplied in the signed SAML assertion. The client then creates a proxy request (includes SAML assertion) and signs the request with the newly created private key. When Dorian receives the proxy request from the client it uses the public key included in the SAML assertion to verify that the proxy request was signed with the clients private key, thus ensuring that the SAML assertion originated from the client that the IdP issued it to.

3.3. Dorian Identity Provider

The Dorian Identity Provider (DorianIdP) gives developers, smaller groups, research labs, unaffiliated users, and other groups that don't have their own IdP, the ability to leverage Dorian. The DorianIdP provides a method for prospective users to register for an account. The DorianIdP can be configured to automatically approve registration requests or can be configured to require an administrative approval. When users register they create a user id and password which they can subsequently use to authenticate with the Dorian IdP. When a user authenticates, the Dorian IdP provides the user with a SAML authentication assertion, which can then be used to authenticate with Dorian's IFS to create grid credentials. The DorianIdP provides mechanisms for administrators to

manage users; this includes modifying user information (name, address, email, etc.), changing passwords, granting and revoking access, and other administrative actions. All operations provided by the Dorian IdP are made available through Dorian's grid service interface. Administrative operations require administrators to authenticate with a trusted grid proxy.

4. Related Work

Few middleware systems have been developed to facilitate the creation, management, and federation of user credentials in the grid. One such system is the MyProxy Credential Management Service [10,11]. MyProxy is an open source project for managing X509 Public Key Infrastructure. MyProxy provides the ability to manage private keys and certificates for grid users. Similar to Dorian it enables users to supply a password to securely create grid proxies based on their private key and certificate stored in the MyProxy repository. When GUMS, the predecessor to Dorian, was developed MyProxy did not have a built in certificate authority and it required its users to upload their private key and certificate. This was a major factor in the decision to develop GUMS, as the built in certificate authority greatly reduces the complexity of the creation grid user credentials. Since that time, MyProxy (MyProxy 3.0) has added the ability to act as a Certificate Authority. The main difference between MyProxy and Dorian is Dorian's support for web service interfaces and its ability to federate institutions' existing users to the grid.

The Portal-based User Registration System (PURSE) [12], provides a friendly interface for users of web applications to register for and obtain access to their grid credentials. PURSE uses SimpleCA and MyProxy for the creation and management of grid credentials (Developed before MyProxy had a built in CA). PURSE differs from Dorian as it is purely web based and more of a toolkit used for simplifying the development of web applications whereas Dorian is a free-standing grid service focused on solving the identity mananagment and federation problem.

5. Conclusion

In a large scale grid, the process of provisioning grid credentials is an important yet difficult task. The task becomes even more complicated when the multiple institutions involved, have disparate user management solutions. The work presented here aims to satisfy the seemingly conflicting requirements of enabling institutions to leverage existing user management solutions while fostering a uniform approach to grid identity. By leveraging (1) existing grid security infrastructure, (2) community standards for security assertions, and (3) the flexibility of institutional user management systems, Dorian provides a grid-managed solution to the problem of identity management and federation for large multi-institutional grids.

6. References

[1] Cancer Biomedical Informatics Grid (caBIG™), https://cabig.nci.nih.gov.

[2] Cancer Biomedical Informatics Grid (caBIG™), https://cabig.nci.nih.gov/workspaces/Architecture/caGrid/

[3] I. Foster, C. Kesselman, S. Tuecke., "The Anatomy of the Grid: Enabling Scalable Virtual Organizations", *International J. Supercomputer Applications*, 15(3), 2001.

[4] I. Foster, C. Kesselman, J. Nick, S. Tuecke, "The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration", 2002.

[5] The Globus Toolkit, http://www.globus.org

[6] V. Welch, F. Siebenlist, I. Foster, J. Bresnahan, K. Czajkowski, J. Gawor, C. Kesselman, S. Meder, L. Pearlman, S. Tuecke, "Security for Grid Services", *Twelfth International Symposium on High Performance Distributed Computing (HPDC-12)*, IEEE Press, to appear June 2003.

[7] I. Foster, C. Kesselman, S. Tuecke, J. Volmer, V. Welch, R. Butler, D. Engert, "A National-Scale Authentication Infrastructure.", *IEEE Computer*, 33(12):60-66, 2000.

[8] S. Langella, S. Oster, S. Hastings, T. Kurc, J. Saltz, "caGrid 0.5 Security Infrastructure". July 2005.

[9] OASIS Security Services (SAML) TC, http://www.oasisopen.org/committees/tc_home.php?wg_abbrev=security

[10] J. Novotny, S. Tuecke, and V. Welch., "An Online Credential Repository for the Grid: MyProxy". Proceedings of the Tenth International Symposium on High Performance Distributed Computing (HPDC-10), IEEE Press, August 2001, pages 104-111.

[11] J. Basney, M. Humphrey, and V. Welch., "The MyProxy Online Credential Repository". Software: Practice and Experience, Volume 35, Issue 9, July 2005

[12] PURSe: Portal-Based User Registration Service, http://www.grids-center.org/solutions/purse/

[13] Ken Lin, Gary Daemer, "caBIG™ Security Technology Evaluation White Paper". January 2006.

[14] Web Services Resource Framework (WSRF), http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=wsrf

[15] S. Tuecke, V. Welch, D. Engert, L. Pearlman, M. Thompson, Internet X.509 Public Key Infrastructure (PKI) Proxy Certificate Profile